유전체 역학연구의 동향
Genomic Epidemiology

Daehee Kang
Seoul National University
College of Medicine

Lecture Outline

- Molecular & Genomic Epidemiology
- Case-Control Study of Genetic Susceptibility
  - SNP Analysis
  - SNPs and Breast Cancer
  - GSEC
- Cohort Study for Genomic Epi Studies
- Future Directions

J Chron Dis Vol. 35, pp. 581 to 600. 1982
Printed in Great Britain

0021-9681/82/070581-20$03.00/0
Pergamon Press Ltd

MOLECULAR EPIDEMIOLOGY AND CARCINOGEN-DNA ADDUCT DETECTION: NEW APPROACHES TO STUDIES OF HUMAN CANCER CAUSATION

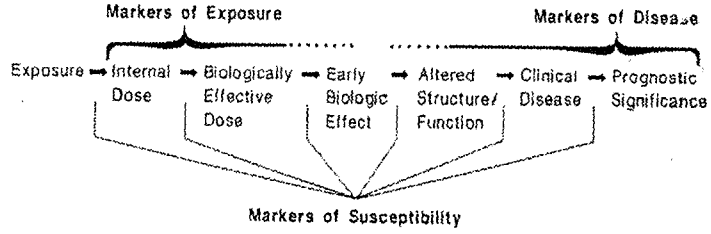
FREDERICA P. PERERA and I. BERNARD WEINSTEIN
Division of Environmental Sciences, School of Public Health and Cancer Center/Institute of Cancer Research, Columbia University, New York City, NY 10032, U.S.A.

(Received 2 July 1981)

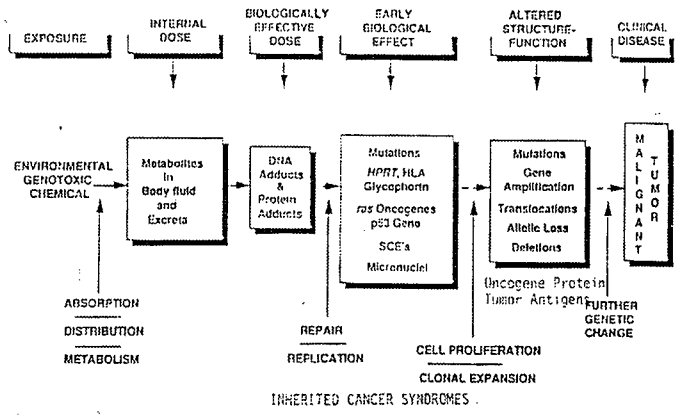
Traditional Epidemiology

Exposure → ■ → Disease

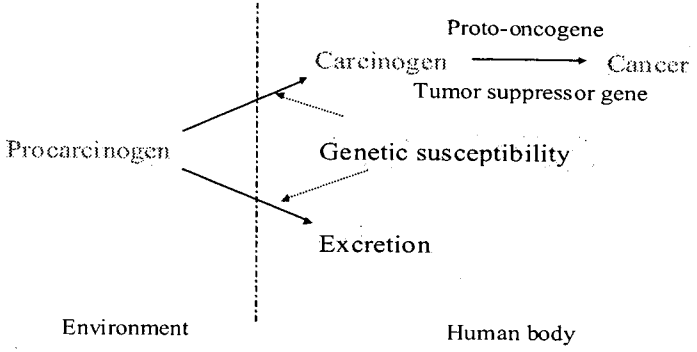
Molecular Epidemiology



MOLECULAR BIOMARKERS OF GENOTOXIC EXPOSURE



Cancer & Genetic Susceptibility



Types of Susceptibility Genes

Which SNPs be analyzed?

- biological relevance
  - previous studies with other diseases
  - haplotype: intron, promotor regions, cSNP, etc.
  - epistasis: up- or down-stream
- allele frequencies
- available genotyping methods

SNP Database

- browse target SNPs

http://snp.cshl.org/

http://www.ncbi.nlm.nih.gov/

http://www.ncbi.nlm.nih.gov/SNP/

http://snp500cancer.nci.nih.gov/home.cfm

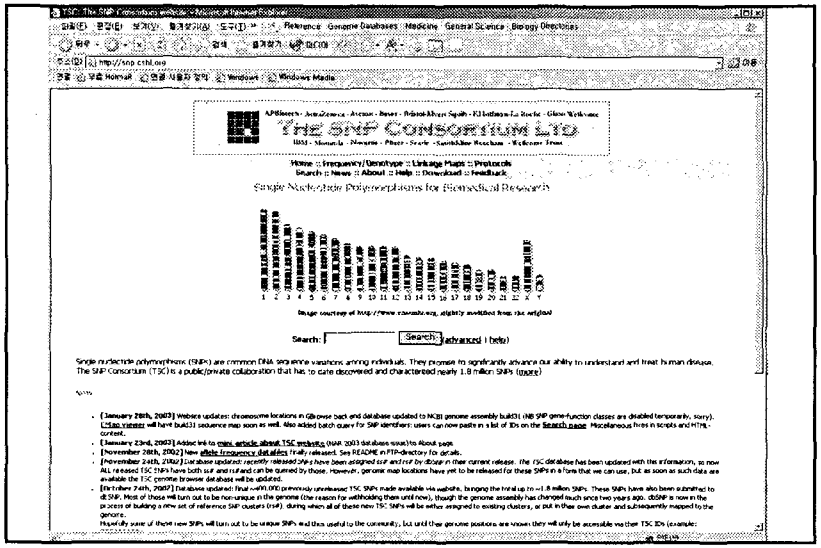
http://www.ncbi.nlm.nih.gov/omim/

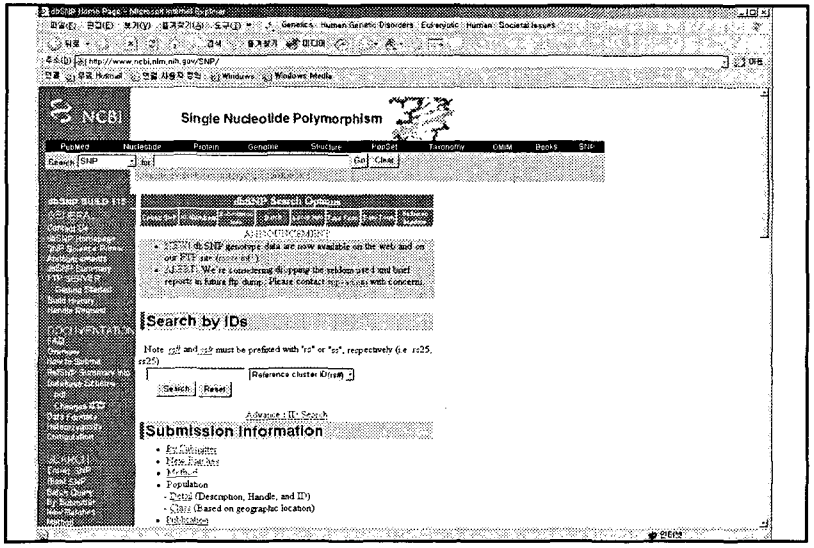
http://lpg.nci.nih.gov/

http://hgvbase.cgb.ki.se/

http://ariel.ucs.unimelb.edu.au:80/~cotton/mdi.htm

http://www2.ebi.ac.uk/mutations/





Ideal genotyping method

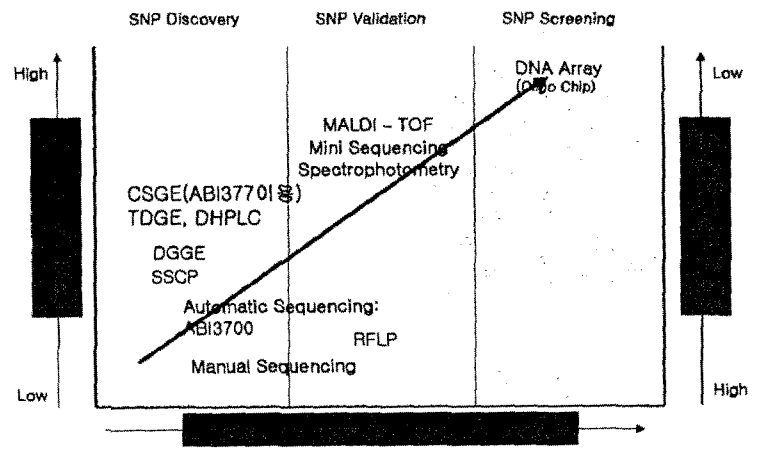
Should be …

- Easily and quickly developed from sequence information
- Low cost for assay development in terms of marker-specific reagents and time spent by expert personnel on optimization
- Robust reaction even at suboptimal DNA samples
- Fully automated procedures and minimal hands-on operation
- Simple, automated, accurate genotype calling for data analysis
- Flexible and scalable assay format (performing a few hundred to a million assays per day)
- Low assay cost for optimized genotyping step (including equipment, reagents, and personnel)

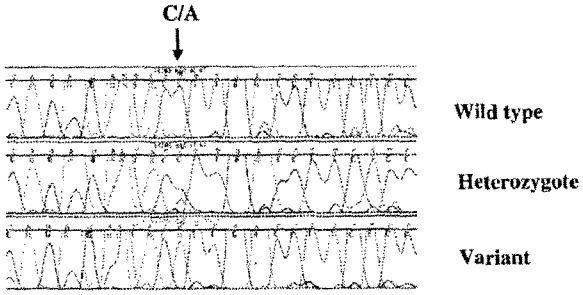
Methods in dbSNP at NCBI

Computed	variation was mined from sequence alignment with software
dHPLC	Denaturing High Pressure Liquid Chromatography used to detect SNP
Hybridize	hybridization method (e.g. chip) was used to assay for variation
RFLP	variation in enzyme restriction site used to detect variation
Sequence	samples were sequenced and resulting alignment used to define variation
SSCP	single stranded conformational polymorphism used to detect variation
Other	other method used to detect variation

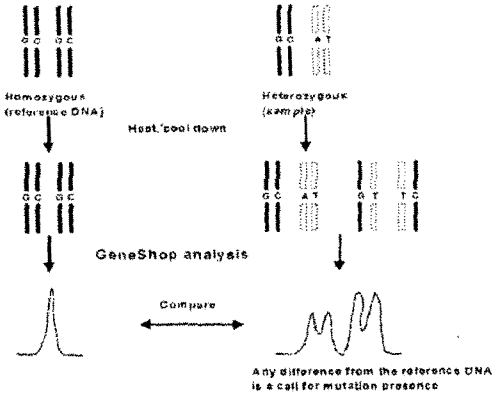
SNP Analysis & Technology



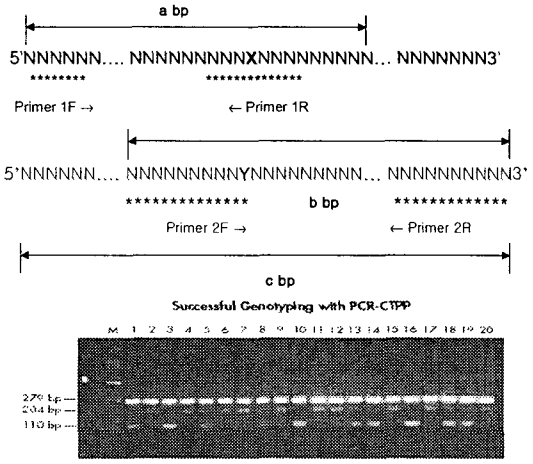
Direct sequencing for SNP discovery



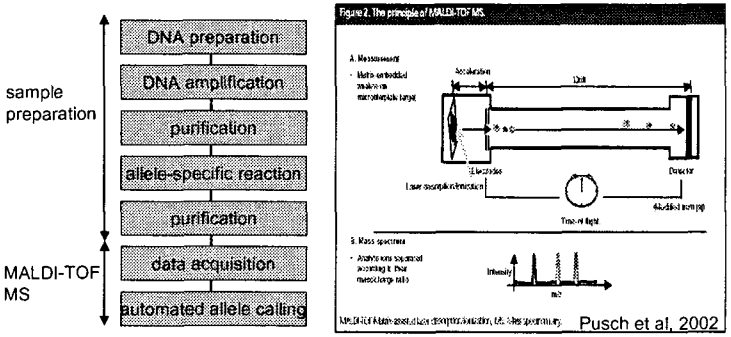
dHPLC

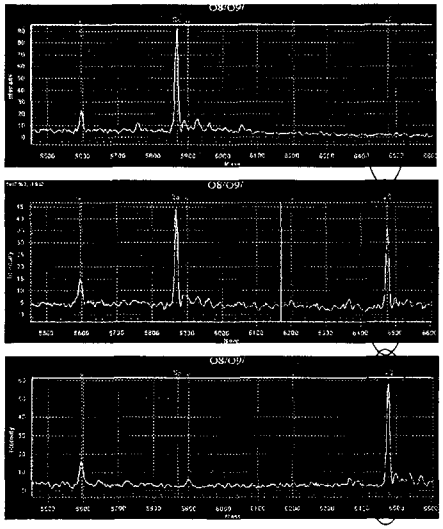


PCR-CTPP

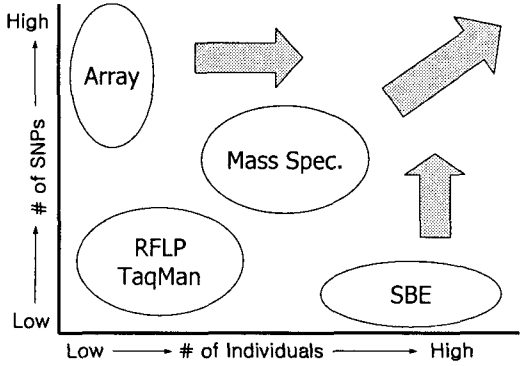


Matrix-assisted laser desorption/ionization (MALDI-TOF) mass spectrometry (MS)





Technology Platform Extension for SNP Screening



SNPs and Breast Cancer

Selected Low Penetrance Genes Studied
MGEL, SNUCM

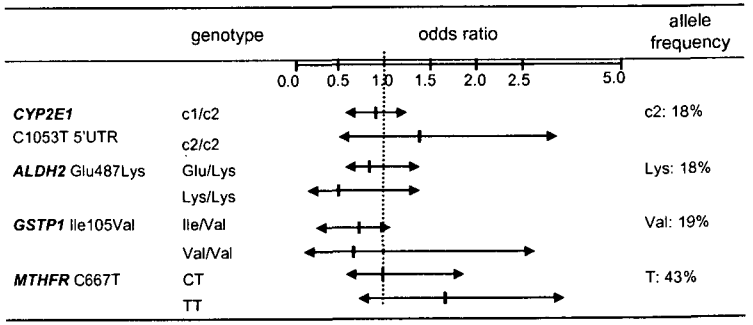
group	genes
xenobiotic metabolism	*CYP1A1/2E1, GSTM1/T1/P1, NAT1/2, ALDH2, EPHX, NQO1,*
estrogen metabolism	*CYP17/19/1B1, COMT, ER-α*
DNA repair	*XRCC1/2/3/4/6, ERCC2/4, ATM, AGT, LIG4, RAD51/52, hOGG1*
cytokine & growth factor	*HER2, TGFβ1, TNFβ, IGF1, IL-1β, IL-RN,*
cell cycle control	*CCND1, CDK7, Bcl-6, MTHFR*

published/in press/revised

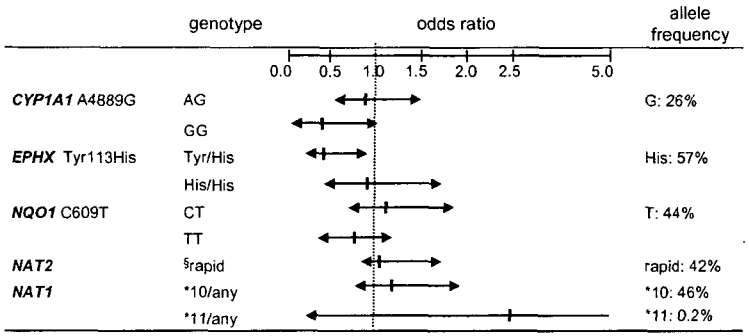
Genotyping Method

Method	*Genes*	No	YR
PCR	*GSTM1/T1*	500	1998
RFLP	*GSTP1, COMT, MTHFR, ERα, XRCC1, hOGG1*		
RT-PCR	*CYP1B1/2E1, NQO1*	700	1999
DASH	*CYP19, ERCC4, EPHX*		
TaqMan	*NAT1/2*		
CTPP	*TGFβ1, TNFβ, IGF1, IL1B, IL-RN, ALDH2, BAR2, HER2*	1100	2000
SnapShot	*CYP1A1/17, ATM, XRCC3, XPD, AGT*	1400	2001
MALDI-TOF	*XRCC2/4/6, LIG4, RAD51/52, CCND1, CDK7, Bcl-6, ERα*	2100	2002

Carcinogen Metabolizing Enzyme Genes and Breast Cancer



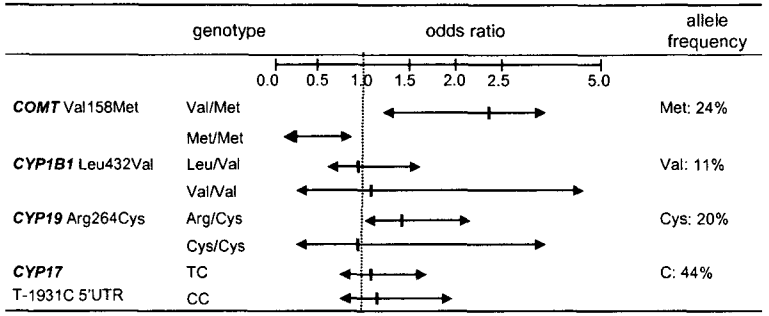
Carcinogen Metabolizing Enzyme Genes and Breast Cancer



§phenotype

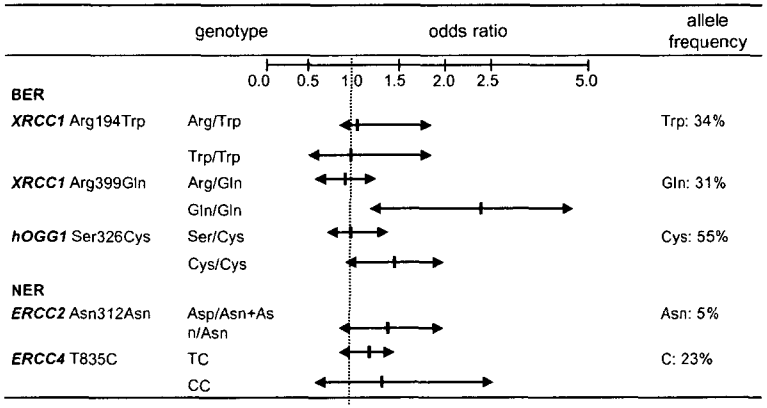
Choi *et al.*, MGE of Cancer-AACR, Hawaii, 2003

Estrogen Metabolizing Enzyme Genes and Breast Cancer Risk



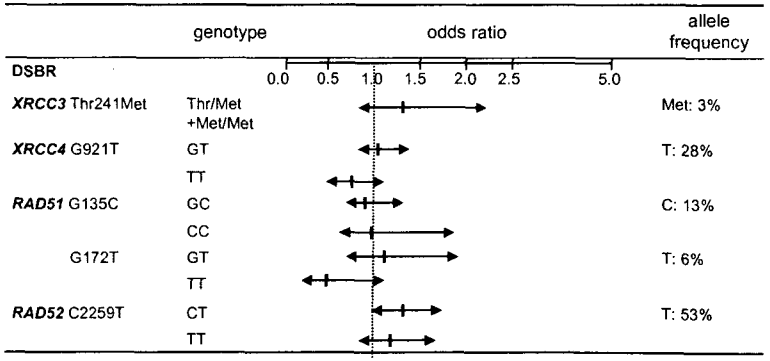
Shin *et al.*, *Proc AACR*, 2003

DNA Repair Genes and Breast Cancer



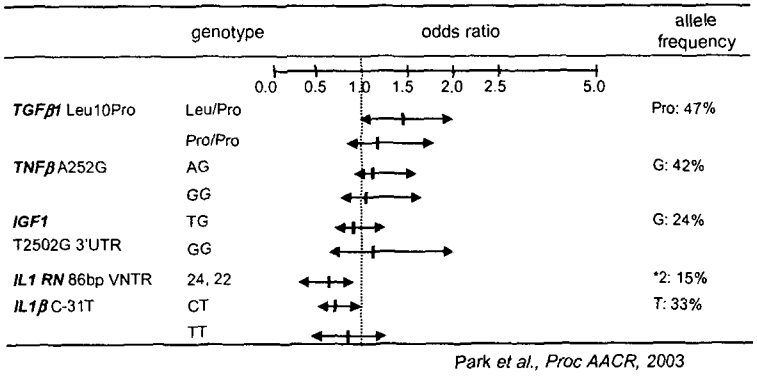
Choi *et al.*, *BCRT*, 2003

DNA Repair Genes and Breast Cancer



Lee *et al.*, *Proc AACR*, 2003

Cytokine and Growth Factor Genes and Breast Cancer



Park et al., Proc AACR, 2003

GSEC (International Collaborative Study on Genetic Susceptibility to Environmental Carcinogen)

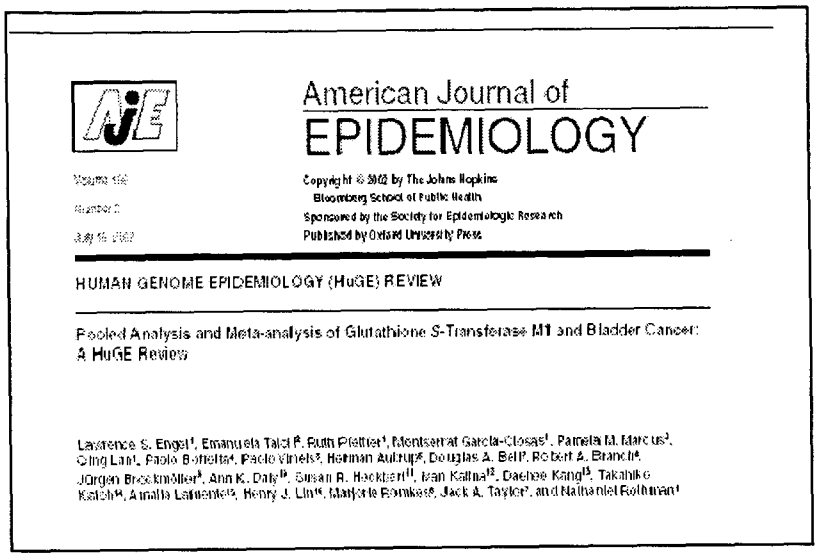
- International Pooled Data Analysis Project -

- Objectives
  - to investigate the association and interaction between polymorphisms in carcinogenesis with adequate statistical power
- Metabolic genetic polymorphisms
  - *CYP1A1, CYP2D6, CYP2E1, GSTM1/T1, NAT2, EH, NAT1, NQO1, GSTP1*
- Current Data Status (Jan 2003)
  - 78 investigators,136 studies
  - total n=44,966 (Case: 21,092, Controls: 23,874)
  - Cancer sites: lung (29%), bladder (15%), breast (14%), head and neck (10%), skin (8%), colon (7%), leukemia (5%), ovary-uterus (3%), prostate (3%), others (7%)

Distribution of Genetic Polymorphisms

Gene	Controls, n (%)	Cases, n (%)
GSTM1	17843 (74.7%)	16937 (80.3%)
GSTT1	10672 (44.7%)	11753 (55.7%)
CYP1A1	10133 (42.4%)	8780 (41.6%)
CYP2E1	4178 (17.5%)	3701 (17.5%)
CYP2D6	5811 (24.3%)	5805 (27.5%)
NAT2	7121 (29.8%)	6108 (29.0%)
GSTM3	1820 (7.6%)	3098 (14.7%)
GSTP1	5166 (21.6%)	5744 (27.2%)
HPX	1144 (4.8%)	1003 (4.8%)
NAT1	1714 (7.2%)	1999 (9.5%)
CYP2C19	1615 (6.8%)	1051 (5.0%)
Two genes	4846 (20.3%)	4962 (23.5%)
Three genes	4270 (17.9%)	3879 (18.4%)
Four genes	3532 (14.8%)	2685 (12.7%)
Five or More genes	4159 (17.4%)	4773 (22.6%)

- Approved projects (in press or to be submitted)
  - Lung cancer, race, histology, smoking, CYP1A1, CYP2E1 *(Le Marchand)*
  - Distribution of GST, smoking and drinking habits among controls *(Benhamou & Smit)*
  - Colorectal cancer risk and GSTM1 deletion *(Smit)*
  - Low dose exposures, residential history and metabolic and metabolic polymorphisms *(Vineis & Veglia)*
  - Genetic susceptibility to head and neck cancer: Gene-gene and gene-environment interactions *(Brennan)*
  - Genetic susceptibility to breast cancer susceptibility to breast cancer and interaction with tobacco smoking *(Boffetta)*



American Journal of
EPIDEMIOLOGY

Volume 156
Number 1
July 1, 2002

Sponsored by the Society for Epidemiologic Research
Published by Oxford University Press

HUMAN GENOME EPIDEMIOLOGY (HuGE) REVIEW

Pooled Analysis and Meta-analysis of Glutathione *S*-Transferase M1 and Bladder Cancer: A HuGE Review

Lawrence S. Engel[1], Emanuela Taioli[2], Ruth Pfeiffer[1], Montserrat Garcia-Closas[1], Pamela M. Marcus[3], Qing Lan[1], Paolo Boffetta[4], Paolo Vineis[5], Herman Autrup[6], Douglas A. Bell[7], Robert A. Branch[8], Jürgen Brockmöller[9], Ann K. Daly[10], Susan R. Hecktert[11], Ivan Kalina[12], Daehee Kang[13], Takahiko Katoh[14], Arnaldo Lauriente[15], Henry J. Lin[16], Marjorie Romkes[6], Jack A. Taylor[7], and Nathaniel Rothman[1]

Cohorts for Genomic Epi

- US NCI Cohort Consortium
- COGENE
- UK Biobank
- EPIC
- Japanese Cohorts: JPHC, JACC
- Korean Cohorts: KMCC

No.	Cohorts	Year began	No. at entry	Female %	Ethnicity
1	AARP Diet and Health Study (US)	1995	540,833	40%	White (92%)
2	Agricultural Health Study (US)	1994	89,190	35%	White (97%)
3	The Black Women's Health Study (US)	1995	64,500	100%	Af-Am (100%)
4	California Teachers Study (US)	1995	133,479	100%	White (87%)
5	Physicians' Health Study I/II (US)	1981/1997	29,571	0%	White (94%)
6	Clue Cohort (US)	1974/1989	58,000	55%	White (98%)
7	Women's Health Study (WHS) (US)	1992	39,876	100%	White (94%)
8	PLCO (US)	1993	148,000	50%	White (90%)
9	Nutrition Intervention Trials in Linxian (China)	1985	32,902	55%	Asian (100%)
10	Multiethnic Cohort (US)	1993	215,251	55%	White (23%)
11	The Shanghai Male Cohort Study (China)	1986	18,244	0%	Asian (100%)
12	The Singapore Chinese Health Study (Singapore)	1993	63,257	56%	Asian (100%)
13	Vitamins & Lifestyle Study (VITAL) (US)	2000	75,000	53%	White (93%)
14	Shanghai Women's Health Study (China)	1992	75,000	100%	Asian (100%)
15	Prevention Study (CPS) II Nutrition Cohort (US)	1982	184,000	53%	White (97%)
16	Health Professionals Follow-up Study (US)	1986	51,529	0%	White (97%)
17	Nurses' Health Study (US)	1976	121,700	100%	White (85%)
18	Nurses' Health Study II (US)	1989	116,671	100%	White (94%)
19	EPIC (Europe 8 countries)	1993	480,000	69%	White
20	ATBC Study (US)	1985	29,133	0%	White (100%)
21	New York Womens' Health Study (US)	1985	14,275	100%	White (85%)

No.	Biospecimen, n (%)	Types of Biospecimen	No. Cancer (by 2002), n (%)
1	270000 (50%) (planning)	buccal cell (planning)	30,000 (6%)
2		buccal cell	2000 (2%)
3		buccal cell	700 (1%)
4	432 (0.3%)	buccal cell cheek brushes	13,400 (10%)
5	I: 15,000 (68%), II: 6150 (82%)	plasma & red cells/buffy	3,607 (12%)
6	58,000 (100%)	serum, plasma, RBC, buffy coat	3,263 (6%)
7	28,133 (71%)	plasma, buffy coat, RBC	1,699 (4%)
8	60,000 (41%)	plasma, buffy coat, RBC	3,000 (2%)
9	32,902 (100%)	plasma, buffy coat, RBC, toenails	3,656 (11%)
10	8,900 (4%)	serum, plasma, RBC, buffy coat, urine	18,300 (9%)
11	18,244 (6%)	serum, urine	1,178 (6%)
12	1,500 (2%)	plasma, serum, RBC, buffy coat, urine	2,355 (4%)
13	35,000 (47%)	buccal cell	N/A
14	60,000 (80%)	plasma, RBC, WBC	1,200 (2%)
15	40,000 (22%)	serum, plasma, buffy coat, RBC	20,000 (11%)
16	33,000 (64%)	plasma, WBC, RBC, buccal cell	4,000 (8%)
17	32,826 (30%)	plasma, buffy coat, RBC	
18	29,500 (25%)	plasma, buffy coat, RBC, urine	
19	400,000 (83%)	serum, plasma, buffy coat, RBC	15,200 (3%)
20	29,133 (100%)	serum, RBC, toenail	7,000 (24%)
21	14,250 (100%)	serum, blood clots	2,000 (14%)

COGENE (Co-ordination of Genomes Research Across Europe)

The COGENE project consists of five work packages:

- creation of a European web site on human genomics
- survey of the current status of genomes research in the European Union.
- informing the public.
- workshop on population genetics
- workshop on pharmacogenomics

Major European Population Cohorts/Biobanks Collected for Genetic Studies (>1000 samples)

Country	Cohorts/Registry	DNA Biobanks/Subjects
Austria	Austrian Disease and Health Bank Project	Frozen tissues: >20 000; Paraffin embedded tissues >2,5 M; Patient data >600 000
Cyprus	Histocompatibility Laboratory at the Paraskevaidion Transplant Center	Approx. 3000 DNA samples
Czech	Masaryk Memorial Cancer Institute	Breast ca. pts & healthy; >1000 DNA samples
	Cystic Fibrosis Registry	>2500 DNA samples
	Thrombofilia Registry	>1000 DNA samples
Estonia	Estonian Genome Project	10 000 participants (DNA & health records)
	Cancer Register	30 000 samples
Finland	Finnish Twin Cohort	170 000 samples
	Health 2000 Cohort	11 500 samples
	Northern Finland Cohorts	21 000 samples
	FINRISK 92	8000 samples (cardiovascular diseases)
	FINRISK 97	10 000 samples (cardiovascular diseases)
	FINRISK 02	11 000 sample (cardiovascular diseases)

Country	Cohorts/Registry	DNA Biobanks/Subjects
Finland	Diabetes Cohort	6000 samples
	Autism Cohort	2000 samples
	Psychosis Cohort	3000 samples
Germany	KORA/MONICA	>20.000 individuals (for cardiovascular disease & diabetes)
	EPICS	>30.000 individuals
	PopGen	2.3 million individuals (for 12 "civilization diseases)
Israel	Ashkenazi Jewish population	several thousands samples (IDgene Pharmaceuticals Ltd)
	The Israel Ischaemic Heart Disease	10,000 civil servant men
	Jerusalem Lipid Research Clinic	7000 middle aged and 8600 offspring
	Jerusalem Perinatal Study	92,000 children & 40,000 of their mothers.
Latvia	Latvian Diabetes Registry	31 000 patients
Netherlands	Cohorts of elderly subjects	from many studies
	Twin Cohorts	Netherlands Twin Register (adult/young twins): depression, cardiovascular etc.
Norway	CONOR – Cohort of Norway	175 000 participants
	MOBA – Norwegian mother and child cohort	16 000 mothers

Country	Cohorts/Registry	DNA Biobanks/Subjects
Portugal	Bleeding disorder register	2800 samples
	Cardiovascular disease register	1746 samples
	Cystic fibrosis register	4100 samples
Sweden	The Västerbotten Project Cohort	70 000 samples
	The Northern Sweden Monica Cohort	7 000 samples
	The Västerbotten Mammary Screening Cohort	35 000 samples
UK	Biobank UK	500 000 participants (planning)
	MRC - Genetic Materials Collection	14 collections of genetic material (complex disease patients)
	National Child Development Study	17 000 participants
	EPIC-Norfolk: Prospective study on cancer	30 000 aged 45-74 yrs
	ALSPAC	14 000 children born in Bristol area 1991-2
	Million Women Study	Breast cancer screening programme
	ProtectT Study	Prostate cancer study

UK Biobank project (2002-9)

-a study on the genes, environment and health-

- large population-based prospective study aimed at establishing how genes, lifestyle and environmental factors interact to affect people's health in UK (2004~)
- fund: MRC (Medical Research Council), Wellcome Trust biomedical charity, and Department of Health
- background
  - small sample size (inconsistent and unreliable results)
  - incomplete or inadequate measures of exposure
  - lack of formal statistical testing
  - retrospective case-control design
  - characteristics of UK
    - a large heterogeneous population
    - Centralized National Health Service

- Objectives
  - to investigate the separate and combined effects of genetic and environmental factors (lifestyle, physiological and environmental exposures) on the risk of common multifactorial diseases of adult life
- Study Design
  - prospective cohort study
  - nested case-control study
- Subjects recruitment
  - ~500,000 (45-69 yrs) (representative sampling)
  - 500~600 general practices
- Data collection
  - self-administered questionnaire: SES, demography, habits/lifestyle, diet (food frequency questionnaire), reproductive history, family history etc.
  - interview on their medical histories, and written consent for participation and follow-up
  - physical assessment : blood pressure, anthropometric (height, weight, waist and hip), FEV1
  - 7-day diet diary

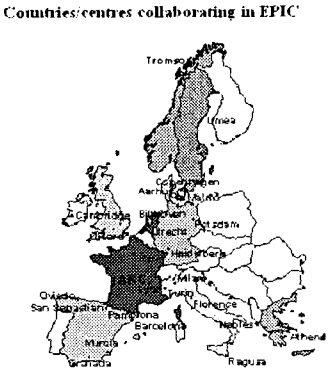
- Blood sampling (50ml)
  - plasma, buffy coat, and whole blood are stored in LN_2
  - various analytes are to be measured
- Follow-up (health monitoring for >10 yrs)
  - many disorders, including cancer, heart disease, diabetes and Alzheimer's disease etc
  - cause-specific mortality and cancer incidence: Office of National Statistics:
  - incident morbidity: hospitalization data, general practice records
  - NHS information system
- Resurvey (exposure update & self-reported incident morbidity)
  - on 2,000 participants in every 2 years
  - on entire cohorts by postal questionnaire

- Analysis
  - series of nested-case control studies
- Statistical power (if, 95% power, 0.1% significance)
  - >5,000 incident cases (e.g., diabetes mellitus, ischaemic heart disease mortality, myocardial infarction, colorectal cancer, breast cancer)
    - minimum relative risk (prevalence of exposure or genotypes 20~80%): 1.5
    - minimum interaction ratio: 1.4
  - 1000-2000 incident cases (e.g., rheumatoid arthritis, Parkinson's disease, hip fracture, ovarian cancer, bladder cancer etc.)
    - minimum relative risk (prevalence of exposure or genotypes 20~80%): 1.8-2.0
    - minimum interaction ratio: 1.7-2.0

참고자료: http://www.biobank.ac.uk/documents/draft_protocol.pdf

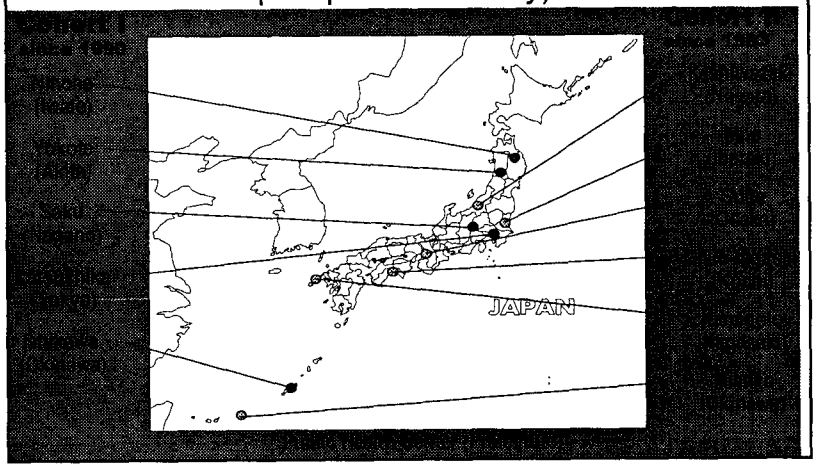
EPIC (1992~)
-European Prospective Investigation into Cancer and Nutrition-

- Objectives: to investigate the relationships between diet, nutritional status, lifestyle and evironmental factors and the incidence of cancer and other chronic diseases
- Fund: Europe Against Cancer Program of the European Commission
- Subjects: >520,000 (≥20 yrs), ten European countries (France, UK, Germany, Greece, Italy, Netherlands, Spain; Sweden Denmark, Norway(1995~))



- Data collection
  - diet and lifestyle: questionnaire
- Bio-bank and biological samples
  - from virtually 100% of participants (except in France and Oxford 33%)
  - 30ml blood -> 0.5ml cryo-plastic straws
  - storage
    - LN_2 (-196°C)
    - local center and IARC
  - analysis
    - Vit C etc.
    - markers of diet related factors
- Follow-up: >10 yrs
- First results were presented in June 2001.

JPHC (Japan Public Health Center-based prospective Study)



- Objectives
  - a population based cohort on diet, cancer and cardiovascular diseases
- Study Design
  - prospective cohort study, nested case-control study
- Subjects (n=141,000)
  - Cohort I (1990~): 62,000 (40~59 yrs)
  - Cohort II (1993~): 79,000 (40~69 yrs)
- Data collection
  - self-administered questionnaire: past medical history, smoking and drinking, diet, family history, physical activity, stress and social support, residential history, occupational history, personality, reproductive history
  - Food Frequency Questionnaire: 44~46 items, 4~5 frequency categories, without portion size
- Blood sampling along with Health Check-up
  - 10ml peripheral blood in heparinized tube
  - centrifuged within 12 hours: 3 tubes (1ml) for plasma, 1 tube (1ml) for buffy layer
  - stored at -80°C
- Follow-up
  - Population registry, death certificate, medical record
  - with 5- or 10-year follow-up survey

Japan Collaborative Cohort (JACC) Study

Baseline study	Follow-up study(-1999)	Organ-specific analysis
Participated areas : 29	Average period:10.1 Years	Stomach and Liver cancer
Collected in 1988-90	Total deaths: 12,178	Large intestinal cancer
Subjects: 110,792	(Cancer:37%)	Pancreas and Gallbladder cancer
Serum samples: 40,000	Cancer incidences: 5,228	Urinary tract and Prostate cancer
Male: 42%, Age: 40-79Ys	Final follow-up (2000/12/31)	Lung cancer

Lifestyle and lung cancer death (Lifestyle analysis)

- Risk factors (RR)
  - Habitual smoking (Males:4.5, Female:3.5)-- Lower than those in the US & UK
  - High intake of fatty foods in non-smokers and obesity at age 20 (females)
- Protective factors
  - Stop smoking (the risk was lowered around 1.0 after 15-20 years)
  - Frequent intake of green-yellow vegetables and fruit in smokers

Serum components and lung cancer death (Serum analysis)

- Positive association: IGF-I (adjusted by IGFBP-3)
- Negative association: IGF-II、IGFBP-3、Total cholesterol, α-carotene, β-Carotene, β-Cryptoxanthin

Multi-institutional Molecular Epidemiology Cohort in Japan (2003~)

- Potential models
  - Hospital-based research (e.g. research program in Aichi Cancer Center)
  - Community-based health control programs
  - Private health screening programs
  - Occupational health screening programs
  - Related cohorts for other study purposes
  - Other specific communities

Molecular Epidemiology Cohort in Hiroshima, Japan (2003~)

-Pilot Study Prior to the Nation-Wide Project-

- Subjects
  - n=7000 (≥40 yrs)
  - 1 town → other towns in 2nd or 3rd year
  - informed consent based on sufficient understanding
- Cooperation with
  - local administration
  - medical institutions
  - Hiroshima University
- Interactive cohort study design
  - lifestyle guidance,
  - measurement of biomarkers in site
  - periodic re-survey

Korean Multi-center Cancer Cohort

(since 1993)

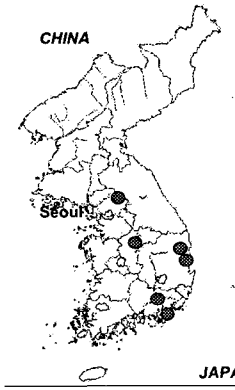
KMCC

1993-2003
6 sites : urban & rural

Haman(1993)
Choongju(1996)
Uljin(1994)
Pohang(1994)
Euryung(2001)
Chooncheon(2003)

questionnaires = 15,983

73,950 person-years



www.kmcc.or.kr

Baseline Information : KMCC (since 1993)

- questionnaire on lifestyles : direct interview
- anthropometric measurements
- clinical laboratory tests : b-chemistry, lipids, stools, tumor markers

Biologic Materials Bank : KMCC (n = 15,641)

blood (20ml)	plasma	*(0.5ml x 3 E-tubes)*
	WBC buffy coat	*(0.5ml x 1 E-tubes)*
	RBC clots	*(0.5ml x 2 E-tubes)*
		2 sets stored at -70°C
urine (50ml)	*(10ml x 2 f-tubes)*	
		2 sets stored at -20°C

Cohort Studies with Blood Samples
(selected from Rothman et al. 2002)

Studies	Year began	No. of subjects with blood samples
NYU Women's Health	**1985**	**14,000**
Nurses' Health, USA	**1989**	**33,000**
Washington County	**1989**	**33,000**
Women's Health Initiative	**1993**	**164,000**
EPIC, Europe	**1993**	**350,000**
Melbourne Collaborative Cohort	1990	42,000
JPHC, Japan	1990	49,000
JACC, Japan	1989	50,000
KMCC, Korea	1993	15,000
Shanghai Women's Health	1997	55,000

Future Directions

- large cohorts for genomic research
  - at least 100,000 enrollments
  - G-E/G-G interaction: etiology/survival
  - Asian consortium for genomic study
- the genotype-phenotype relationship
  - expression and/or functional assay
- application to
  - pharmacogenomics: individual variation of drug response
  - preventive trial